

**GenePattern: A Software Environment for Microarray Analysis****Reich, M.<sup>\*</sup>, Lerner, J., Ohm K., Tamayo, P., Mesirov, J.P.****Whitehead/Massachusetts Institute of Technology Center for Genome Research, Cambridge, MA, USA**

Research in the biological sciences is becoming increasingly dependent upon computational work to simulate biological systems and to analyze the increasing amounts of data generated by high throughput laboratory technologies. There is, however, an inherent tension in providing software solutions that furnish both a packaged solution for the bench scientist and a sandbox for the computational biologist. The latter does not want to incur a lot of software engineering overhead to integrate, evaluate, and deploy a new method. In an active research environment new computational methods are constantly being developed and tested and the ability to integrate new methods with existing preprocessing tools, visualizations, and analytic approaches can facilitate their evaluation and push forward specific research goals and agenda. An additional challenge is the increasing complexity and multi-step nature of many analysis methodologies where a typical problem-specific method involves computation and data access across a multitude of tools, platforms, and data repositories.

To address these challenges for microarray analysis, we have developed GenePattern, a system that provides an open source, platform-independent, flexible and extensible software infrastructure to develop, explore, interoperate, and deploy prototypes, analysis algorithms, and methodologies. GenePattern provides a web services-based infrastructure to enable the platform-independent agnostic integration of independently developed microarray analysis modules using a self-service model with no programming effort. It also contains a sophisticated, multi-access extensible *user environment* including: a language and scripting front end (based on the *R* language), a high-level graphical user interface, and an analytic pipeline builder, that allows a naïve user to build and manage complex computational multi-module methodologies. GenePattern is designed to concurrently support, a less computationally sophisticated user community, sophisticated computational methodology development, and quick deployment of new methods.

GenePattern is populated with a core set of microarray analysis modules including pattern recognition, classification and functional annotation. We plan to extend this environment to cover a broader range of computational biology applications including sequence analysis, chemical genomics, proteomics, etc.

*Federal grant support: R33 CA97556-01 (Mesirov) GenePattern: Methodologies and Software Tools for Molecular Pattern Recognition and Discovery.*